



UNITED STATES PATENT AND TRADEMARK OFFICE

I, Susan ANTHONY BA, ACIS,

Director of RWS Group plc, of Europa House, Marsham Way, Gerrards Cross, Buckinghamshire, England declare;

1. That I am a citizen of the United Kingdom of Great Britain and Northern Ireland.
2. That the translator responsible for the attached translation is well acquainted with the French and English languages.
3. That the attached is, to the best of RWS Group plc knowledge and belief, a true translation into the English language of the accompanying copy of the specification filed with the application for a patent in France on 3 September 1999 under the number 99/11,097 and the official certificate attached hereto.
4. That I believe that all statements made herein of my own knowledge are true and that all statements made on information and belief are true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the patent application in the United States of America or any patent issuing thereon.

For and on behalf of RWS Group plc

The 9th day of March 2004



---

# P A T E N T

---

## UTILITY CERTIFICATE – CERTIFICATE OF ADDITION

### OFFICIAL COPY

The Director-General of the Institut National de la Propriété Industrielle certifies that the attached document is a true copy of an application for industrial property titleright filed at the Institute.

Drawn up in Paris, 29 AUGUST 2000

On behalf of the Director-General of the  
Institut National de la Propriété Industrielle  
The Patent Department Head

[signature]

Martine PLANCHE

INSTITUT  
NATIONAL DE  
LA PROPRIÉTÉ  
INDUSTRIELLE

REGISTERED OFFICE  
26 bis, rue de Saint Petersburg  
75800 PARIS Cédex 08  
Telephone: 01 53 04 53 04  
Fax: 01 42 93 59 30

**INPI**INSTITUT NATIONAL DE LA  
PROPRIETE INDUSTRIELLE26 bis, rue de Saint Pétersbourg  
75800 Paris Cedex 08

Telephone: 01 53 04 53 04    Telefax: 01 42 93 59 30

**PATENT, UTILITY CERTIFICATE**

Intellectual Property Code - Book VI

**Cerfa**

No. 55-1328

**REQUEST FOR GRANT**Confirmation of filing by fax ☐

This form is to be completed in black ink and in block capitals

<b>Reserved for the INPI</b>		<b>1. NAME AND ADDRESS OF THE APPLICANT OR THE REPRESENTATIVE TO WHOM THE CORRESPONDENCE IS TO BE ADDRESSED</b>		
DATE OF SUBMISSION OF THE DOCUMENTS	03.09.99	CABINET REGIMBEAU 26, Avenue Kléber 75116 PARIS		
NATIONAL REGISTRATION	99/11,097			
DEPARTMENT OF FILING	75			
DATE OF FILING	03/09/99			
<b>2. APPLICATION</b>		No. of permanent power of attorney	Correspondent's references	Telephone
<input checked="" type="checkbox"/> patent		238012 D18374 MIP      01 45 00 92 02		
<input type="checkbox"/> divisional application				
→ initial application ↓				
<input type="checkbox"/> utility certificate		<input type="checkbox"/> utility certificate No.      date		
<input type="checkbox"/> conversion of a European patent application				
<input type="checkbox"/> patent				
<b>Compilation of the search report</b>		<input type="checkbox"/> deferred <input checked="" type="checkbox"/> immediate		
The applicant, as a physical person, asks to pay the fee by instalments		<input type="checkbox"/> yes <input type="checkbox"/> no		
<b>Title of the invention</b> (maximum 200 characters)				
CLONING, EXPRESSION AND CHARACTERIZATION OF THE SPG4 GENE RESPONSIBLE FOR THE MOST COMMON FORM OF AUTOSOMAL DOMINANT SPASTIC PARAPLEGIA.				
<b>3. APPLICANT(S)</b>		SIREN No.		APE-NAF code
Name and forenames (underline the surname) or company name				
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)				
Nationality/Nationalities French		Legal form		
		PUBLIC ESTABLISHMENT OF SCIENTIFIC AND TECHNO NATURE ...		
<b>Full address(es)</b>		<b>Country</b>		
3, rue Michel Ange, 75016 PARIS		FR		
If insufficient space, continue on plain paper <input type="checkbox"/>				
<b>4. INVENTOR(S)</b>		The inventors are the applicants		If the answer is no, provide a separate designation
		<input type="checkbox"/> yes <input checked="" type="checkbox"/> no		
<b>5. REDUCTION OF THE RATE OF FEES</b>		<input type="checkbox"/> requested for the first time <input type="checkbox"/> requested prior to filing; attach copy of the favourable decision		
<b>6. PRIORITY DECLARATION OR APPLICATION FOR THE BENEFIT OF THE FILING DATE OF A PRIOR APPLICATION</b>				
Country of origin		Number	Filing date	Nature of the application
<b>7. DIVISIONS</b> previous to the present application		No.	date	No.      date
<b>8. SIGNATURE OF THE APPLICANT OR REPRESENTATIVE</b> (name and capacity of the signatory - registration No.)		<b>SIGNATURE OF THE RECEIVING OFFICIAL</b>		<b>SIGNATURE AFTER REGISTRATION OF THE APPLICATION AT THE INPI</b>
921169      (illegible signature)				(illegible signature)

**INPI**

INSTITUT  
NATIONAL DE  
LA PROPRIETE  
INDUSTRIELLE

**PATENT, UTILITY CERTIFICATE**

**DESIGNATION OF THE INVENTOR**

(if the applicant is not the inventor or the sole inventor)

**PATENTS ADMINISTRATIVE DIVISION**

26bis, rue de Saint-Petersbourg  
75800 Paris Cédex 08  
Tel: 01 53 04 53 04 - Fax: 01 42 93 59 30

NATIONAL REGISTRATION NO.

**99/11,097**

**TITLE OF THE INVENTION:**

CLONING, EXPRESSION AND CHARACTERIZATION OF THE SPG4 GENE RESPONSIBLE FOR THE MOST COMMON FORM OF AUTOSOMAL DOMINANT SPASTIC PARAPLEGIA.

**THE UNDERSIGNED**

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)  
3, rue Michel Ange, 75016 PARIS

**DESIGNATE(S) AS INVENTOR(S) (surname underlined, forenames, address):**

WEISSENBACH Jean  
163, rue de Vaugirard  
75015 Paris, FR

HAZAN Jamilé  
52, avenue René Coty  
75014 PARIS, FR

**NOTE:** In exceptional cases, the name of the inventor may be followed by that of the company to which he belongs (membership company) when the latter is other than the company which is the applicant or proprietor.

Date and signature(s) of the applicant(s) or of the representative

3 September 1999  
(illegible signature)

CABINET REGIMBEAU  
921169

**DOCUMENT CONTAINING AMENDMENTS**

(FRENCH) PAGE(S) OF THE DESCRIPTION OR OF THE CLAIMS OR SHEET(S) OF DRAWINGS			R.M.*	DATE OF THE CORRESPONDENCE	DATE STAMP OF THE CORRECTOR
Amended	Omitted	Added			
p. 43				31. 03. 00	04 APR. 2000 - VD
p. 123-127		p. 128-135	x	31. 03. 00	04 APR. 2000 - VD
p. 132			x	10. 04. 00	13 APR. 2000 - VD

\* A change made in the wording of the original claims, unless the change derives from the provisions of Article R.612-36 of the Intellectual Property Code, is indicated by the reference "R.M." (amended claims).



ORIGINAL

1

CLONING, EXPRESSION AND CHARACTERIZATION OF THE SPG4 GENE RESPONSIBLE FOR THE MOST COMMON FORM OF AUTOSOMAL DOMINANT SPASTIC PARAPLEGIA.

5           The invention relates to the identification and characterization of the SPG4 gene encoding spastin, which is responsible for the most common form of autosomal dominant hereditary spastic paraplegia (HSP), to the cloning and characterization of its cDNA, and also to the corresponding polypeptides. The invention also relates to  
10           vectors, to transformed cells and to transgenic animals, and also to diagnostic methods and to methods for selecting a chemical or biochemical compound capable of interacting directly or indirectly with a polypeptide according to the invention.

          Hereditary spastic paraplegias (HSPs) are degenerative disorders of the central nervous system, characterized by bilateral and progressive spasticity of the lower limbs. They reveal themselves clinically through difficulties in walking possibly evolving  
15           into total paralysis of both legs. The physiopathology of this set of diseases is, to date, relatively undocumented; however, anatomopathological data make it possible to conclude that the attack is limited to the pyramidal tracts responsible for voluntary motricity in the spinal cord (1). Various clinical and genetic forms of HSP exist. The so-called "pure" HSPs, which correspond to isolated spasticity of the lower limbs, are  
20           clinically distinguished from the "complex" HSPs, for which the spasticity of the legs is associated with other clinical signs of neurological or non-neurological type (2). From a genetic point of view, the HSPs can be transmitted according to the autosomal dominant (AD-HSP), autosomal recessive (AR-HSP) or X-linked (X-HSP) mode. The "pure" form of HSP, which is most commonly transmitted according to the autosomal  
25           dominant mode, remains the most frequent (approximately 80% of HSPs) (1). The incidence of HSPs, which remains difficult to estimate because of rare epidemiological studies and the considerable clinical variability, varies from 0.9 : 100 000 in Denmark, 3 to 9.6 : 100 000 in certain regions of Spain (4) or 14 : 100 000 in Norway (5) (approximately 3 : 100 000 in France).

chromosome 2 is a main locus of this form of the disease, found in 40 to 50% of the families analyzed (11, 12). An anticipation phenomenon was observed in some locus SPG4-linked HSP families; this phenomenon has, subsequently, been associated with the expansion of a (CAG)<sub>n</sub> repeat demonstrated in 6 Danish families (13) using the  
 5 RED (for Rapid Expansion Detection) technique. It has, however, never been possible to confirm this expansion in any of the families tested by this method or by the systematic search for sequences of (CAG)<sub>n</sub> type in physical maps composed of YAC (for Yeast Artificial Chromosome) or BAC (for Bacterial Artificial Chromosome) clones (Hazan et al., in press Genomics).

10 To date, three genes responsible for two forms of X-HSP and one form of AR-HSP have been identified. Mutations in the gene which encodes a neuron-specific cell adhesion molecule, L1-CAM (for L1 Cell Adhesion Molecule), and which is located at Xq28 (locus SPG1) cause a complex form of HSP (14) in which the spasticity is associated with a mental handicap, whereas mutations in the PLP (for ProteoLipid  
 15 Protein) gene located at Xq21 (locus SPG2), which encodes a constitutive molecule of the myelin layer, cause pure and complex forms of X-HSP (15). More recently, mutations in the gene located at 16q24.3 (locus SPG7), which encodes paraplegin, a mitochondrial ATPase of the AAA (for "ATPases Associated with diverse cellular  
 20 AR-HSP (17) suggesting that alterations to oxidative phosphorylation (OXPHOS) may be the cause of HSP.

Thus, there remains, today, a great need to identify and characterize the gene responsible for the most common form of AD-HSP. The identification of this gene should, in particular, allow, besides the possibility of a test for antenatal screening in  
 25 the families concerned, a better understanding of some of the molecular mechanisms engendering these degenerations specific for nerve bundles of the spinal cord, or even make it possible to provide an elementary response regarding therapeutic treatment for the patients.

This is precisely the subject of the present invention.

30 After having delimited the localization range between the D2S352 and D2S2347 genetic markers by studying recombination events in locus SPG4-linked HSP families, the inventors have established a

G (B763N4), has revealed the presence of a gene which is composed of 17 exons, extending over a distance of approximately 100 kb, and which exhibits homology with the genes encoding proteins of the AAA family. Comparison of the sequence of this gene between the healthy and affected individuals of AD-HSP families has made it possible to demonstrate various mutations in the patients.

A subject of the invention is thus the identification and characterization of the SPG4 (or SPAST) gene encoding a novel nuclear member of the AAA family, responsible for the most common form of AD-HSP.

In a first aspect, a subject of the present invention is a purified or isolated nucleic acid of the SPG4 gene, characterized in that it comprises at least 15 consecutive nucleotides, preferably 20, 25, 30, 50, 100 or 200 consecutive nucleotides, of a sequence chosen from the group comprising:

- the sequence SEQ ID No. 1, which is a genomic sequence of the human SPG4 gene;
- the nucleic acid sequences which are homologues or variants of the nucleic acid of sequence SEQ ID No. 1;
- the sequence which is complementary thereto; and
- the sequence of the corresponding RNA thereof.

The present invention relates, of course, to both the DNA and RNA sequences, and also the sequences which hybridize with them, as well as the corresponding double-stranded DNAs.

The terms "nucleic acid", "nucleic acid sequence" or "sequence of nucleic acid", "polynucleotide", "oligonucleotide", "polynucleotide sequence", and "nucleotide sequence", which will be used equally in the present description, will be intended to refer to both a double-stranded DNA, a single-stranded DNA and products of transcription of said DNAs, and/or an RNA fragment, said isolated natural, or synthetic fragments which may or may not include unnatural nucleotides, referring to a precise series of nucleotides, which may or may not be modified, making it possible to define a fragment or a region of a nucleic acid. The expression "natural isolated, or synthetic DNA and/or RNA fragment, which may or may not include unnatural nucleotides" is intended to mean a precise series of nucleotides, which may or may not be modified, making it possible to define a fragment, a segment or a region of a nucleic acid.

It should be understood that the present invention does not relate to the genomic nucleotide sequences in their natural chromosomal environment

have been removed directly or indirectly, for example by copying, their environment having been at least partially modified.

The term "homologous nucleic acid sequence" is intended to refer to the sequences which have, with respect to the reference nucleic acid sequence, certain  
5 modifications, such as in particular a deletion, a truncation, an extension, a chimeric fusion and/or a mutation, in particular a point mutation, and the nucleic acid sequence of which shows at least 80%, preferably 90% or 95%, identity after alignment, with the reference nucleic acid sequence. It preferably involves sequences for which the complementary sequences are capable of hybridizing specifically with one of the  
10 sequences of the invention. Preferably, the specific or high stringency hybridization conditions will be such that they ensure at least 80%, preferably 90% or 95%, identity after alignment between one of the two sequences and the sequence which is complementary to the other.

Hybridization under high stringency conditions means that the temperature and  
15 ionic strength conditions are chosen such that they allow the hybridization between two complementary DNA fragments to be maintained. By way of illustration, high stringency conditions of the hybridization step for the purposes of defining the polynucleotide fragments described above are advantageously as follows.

The DNA-DNA or DNA-RNA hybridization is carried out in two steps:  
20 (1) prehybridization at 42°C for 3 hours in phosphate buffer (20 mM, pH 7.5) containing 5 x SSC (1 x SSC corresponds to a 0.15 M NaCl + 0.015 M sodium citrate solution), 50% of formamide, 7% of sodium dodecyl sulphate (SDS), 10 x Denhardt's, 5% of dextran sulphate and 1% of salmon sperm DNA; (2) actual hybridization for 20 hours at a temperature dependent on the size of the probe (i.e. 42°C for a probe of size > 100  
25 nucleotides), followed by two 20-minute washes at 20°C in 2 x SSC + 2% SDS and one 20-minute wash at 20°C in 0.1 x SSC + 0.1% SDS. The final wash is carried out in 0.1 x SSC + 0.1% SDS for 30 minutes at 60°C for a probe of size > 100 nucleotides. The high stringency hybridization conditions described above for a polynucleotide of defined size will be adjusted by those skilled in the art for oligonucleotides of greater or  
30 smaller size, according to the teaching of Sambrook et al., 1989.

The term "nucleic acid sequence which is a variant"

polymorphisms present in mammals, in particular in human beings, and in particular to polymorphisms which can cause a pathology to occur and/or to develop.

While the sequences according to the invention relate to normal sequences, they also relate to sequences which are mutated insofar as they include at least one point mutation, and preferably at most 10% of mutations, with respect to the normal sequence.

In particular, the variant nucleic acid sequences will comprise any sequence of at least 15 consecutive nucleotides, preferably 20, 25, 30, 50, 100 or 200 consecutive nucleotides, of a polymorphic sequence of the genomic sequence of the human SPG4 gene of sequence SEQ ID No. 1, and the nucleic acid sequence of which has, with respect to the sequence SEQ ID No. 1, at least one mutation corresponding in particular to a truncation, deletion, substitution and/or addition of an amino acid residue. In the present case, the variant nucleic acid sequences having at least one mutation will herein be linked to the pathologies of AD-HSP type linked to SPG4 locus.

Preferably, the present invention relates to the mutated nucleic acid sequences in which the mutations produce a modification of the amino acid sequence of the polypeptide encoded by the normal sequence.

The term "variant nucleic acid sequences" will also be intended to refer to any RNA or cDNA resulting from a mutation of a splice site of the genomic nucleic acid sequence SEQ ID No. 1.

The invention preferably relates to a purified or isolated nucleic acid according to the present invention, characterized in that it comprises a sequence chosen from the following group:

- the sequence SEQ ID No. 1;
- the sequence SEQ ID No. 2, which is the cDNA sequence encoding human spastin;
- the sequence SEQ ID No. 72, sequence of the incomplete cDNA encoding murine spastin represented in Figure 5, "mouse" line;
- the nucleic acid sequences which are homologues or variants of the sequences SEQ ID No. 1, SEQ ID No. 2 or SEQ ID No. 72;
- the sequence complementary thereto; and
- the sequence of the corresponding RNA thereof.

The primers or probes, characterized in that they comprise a sequence of a nucleic acid according to the invention, also form part of the invention.

5 The present invention thus relates to the set of primers which can be deduced from the nucleotide sequences of the invention and which may make it possible to demonstrate said nucleotide sequences of the invention, in particular the mutated sequences, using in particular an amplification method such as the PCR method, or a related method.

10 The present invention also relates to the set of probes which can be deduced from the nucleotide sequences of the invention, in particular from the sequences capable of hybridizing with them, and which may make it possible to demonstrate said nucleotide sequences, in particular to distinguish the normal sequences from the mutated sequences.

The present invention relates, in particular, to the probes or primers having sequences chosen from the sequences SEQ ID No. 4 to SEQ ID No. 71.

15 The invention also relates to the use of a nucleic acid sequence according to the invention as a probe or primer, for detecting, identifying, assaying or amplifying a nucleic acid sequence.

20 According to the invention, the polynucleotides which can be used as a probe or as a primer in processes for detecting, identifying, assaying or amplifying a nucleic acid sequence will have a minimum size of 15 bases, preferably of 20 bases, or better still of 25 to 30 bases.

25 The set of probes and primers according to the invention may be labelled directly or indirectly with a radioactive or nonradioactive compound, using methods well known to those skilled in the art, in order to obtain a detectable and/or quantifiable signal.

The nonlabelled polynucleotide sequences according to the invention can be used directly as a probe or primer.

30 The sequences are generally labelled so as to obtain sequences which can be used for many applications. The labelling of the primers or of the probes according to the invention is carried out with radioactive elements or with nonradioactive molecules.

35 Among the radioactive isotopes used, mention may be made of  $^{32}\text{P}$ ,  $^{33}\text$

The polynucleotides according to the invention can thus be used as a primer and/or probe in processes using, in particular, the PCR (polymerase chain reaction) technique (Erich, 1989; Innis et al., 1990, and Rolfs et al., 1991). This technique requires choosing pairs of oligonucleotide primers framing the fragment which must be  
5 amplified. Reference may, for example, be made to the technique described in American patent US No. 4,683,202. The amplified fragments can be identified, for example after agarose or polyacrylamide gel electrophoresis, or after a chromatographic technique such as gel filtration or ion exchange chromatography, and then sequenced. The specificity of amplification can be controlled using, as a primer,  
10 the nucleotide sequences of polynucleotides of the invention and, as a matrix, plasmids containing these sequences or the derived amplification products. The amplified nucleotide fragments can be used as reagents in hybridization reactions in order to demonstrate the presence, in a biological sample, of a target nucleic acid having a sequence complementary to that of said amplified nucleotide fragments.

15 The invention is also directed toward the nucleic acids which can be obtained by amplification using primers according to the invention.

Other techniques for amplifying the target nucleic acid can be advantageously employed as an alternative to PCR (PCR-like), using pairs of primers having nucleotide sequences according to the invention. The term "PCR-like" will be intended to refer to  
20 all methods using direct or indirect reproductions of nucleic acid sequences, or in which the labelling systems have been amplified. These techniques are, of course, known. In general, they involve amplifying the DNA with a polymerase; when the sample of origin is an RNA, it is advisable to perform reverse transcription beforehand. There are, currently, a great many processes which enable this amplification, such as for example  
25 the SDA (Strand Displacement Amplification) technique (Walker et al., 1992), the TAS (Transcription-based Amplification System) technique described by Kwoh et al. in 1989, the 3SR (Self-Sustained Sequence Replication) technique described by Guatelli et al. in 1990, the NASBA (Nucleic Acid Sequence Based Amplification) technique described by Kievitis et al. in 1991, the TMA (Transcription Mediated Amplification)  
30 technique, the LCR (Ligase Chain Reaction) technique described by Landegren et al. in 1988 and improved by Barany et al. in 1991, which uses a heat-stable ligase, the RCR (Repair Chain Reaction) technique described by Segev in 1992, the CPR (Cycling Probe Reaction) technique described by Duck et al. in 1990, and the Q-beta-replicase amplification technique described by Miele et al. in

Chu et al. in 1986 and Lizardi et al. in 1988, and then by Burg et al., and also by Stone et al., in 1996.

When the target polynucleotide to be detected is an mRNA, use will advantageously be made, prior to carrying out an amplification reaction using the primers according to the invention or carrying out a detection process using the probes of the invention, of an enzyme of reverse transcriptase type in order to obtain a cDNA from the mRNA contained in the biological sample. The cDNA obtained will then serve as a target for the primers or probes used in the amplification or detection process according to the invention.

The probe hybridization technique can be carried out in diverse ways (Matthews et al., 1988). The most general method consists in immobilizing the nucleic acid extracted from the cells of various tissues or from cells in culture, on a support (such as nitrocellulose, nylon or polystyrene), and in incubating the immobilized target nucleic acid with the probe, under well defined conditions. After hybridization, the excess probe is eliminated and the hybrid molecules formed are detected using the appropriate method (measurement of the radioactivity, of the fluorescence or of the enzymatic activity linked to the probe).

According to another embodiment of the nucleic acid probes according to the invention, the latter can be used as a capture probe. In this case, a probe, termed "capture probe", is immobilized on a support and is used to capture, by specific hybridization, the target nucleic acid obtained from the biological sample to be tested, and the target nucleic acid is then detected using a second probe, termed "detection probe", labelled with an easily detectable element.

The splice acceptor or donor site sequences identified in Table 3 also form part of the present invention.

In another aspect, the invention comprises a method for screening cDNA or genomic DNA libraries, or for cloning isolated genomic or cDNA encoding spastin, characterized in that it uses a nucleic acid sequence according to the invention.

Among these methods, mention may be made in particular of :

- the screening of cDNA libraries and the cloning of the isolated cDNAs (Sambrook et al., 1989; Suggs et al

chromosomal localization, and then the complete sequencing, of the SPG4 gene encoding spastin.

In particular, these methods according to the invention may be used for identifying and thus obtaining the genomic sequence or the cDNA of the SPG4 gene in  
5 other mammals, in particular mice.

These screening and/or cloning methods will comprise, in particular, a step of hybridization of a nucleic acid according to the invention with a nucleic acid contained in a genomic or cDNA library.

The invention also comprises a method for identifying the nucleic acid  
10 sequences which promote and/or regulate the expression of the SPG4 gene of sequence SEQ ID No. 1, characterized in that it uses a nucleic acid according to the invention.

The computer tools available to those skilled in the art enable them to easily identify, using the genomic nucleic acid sequences according to the invention, the  
15 promoter regulatory boxes required and sufficient for controlling gene expression, in particular the TATA, CCAAT and GC boxes, and also the stimulatory regulatory sequences ("enhancers"), or inhibitory regulatory sequences ("silencers"), which control, in CIS, the expression of the genes according to the invention; among these regulatory sequences, mention should be made of IRE, MRE and CRE.

20 The invention also relates to the methods for identifying mutations carried by the human SPG4 gene, in particular mutations responsible for autosomal dominant hereditary spastic paraplegia, characterized in that they use a nucleic acid sequence according to the invention.

These methods for identifying these mutations will, in particular, comprise the  
25 following steps: (i) isolation of the DNA from the biological sample to be analyzed, or production of a cDNA from the mRNA of the biological sample; (ii) specific amplification of the target DNA likely to have a mutation, using primers according to the invention; (iii) analysis of the amplification products, in particular the size and/or the sequence of the amplification products, with respect to a reference sequence.

30 The expression "methods for identifying a mutation according to the invention" is also intended to refer to a method which makes it possible to obtain the nucleic acid on which said mutation has been identified.

The promoter and/or regulatory sequences of the SPG4 gene according to the invention having mutations which may modify the expression of the corresponding  
35 protein also form part of the invention.</

The nucleic acids characterized in that they can be obtained using one of the preceding methods according to the invention, or the nucleic acids capable of hybridizing, under high stringency conditions (homology of at least 80% between one of the two sequences and the sequence complementary to the other), with said nucleic acids, form part of the invention, especially the variant or homologous nucleic acids, in particular the nucleic acid sequences of allelic variants of the SPG4 gene of sequence SEQ ID No. 1 or of its cDNA of sequence SEQ ID No. 2, and also the genomic sequences of the homologous genes of other mammals such as mice.

In the present description, the term "Spg4" will be intended to refer to the mouse gene homologous to the human SPG4 gene.

The use of a nucleic acid sequence according to the invention as a probe or primer for screening a genomic library or a cDNA of course forms part of the subject of the present invention.

In another aspect, the invention comprises a purified or isolated polypeptide encoded by a nucleic acid according to the invention.

In the present description, the term "polypeptide" will be used to refer equally to a protein or a peptide.

Preferably, the present invention relates to a polypeptide, characterized in that it comprises an amino acid sequence chosen from the following group:

- the sequence SEQ ID No. 3, corresponding to human spastin encoded by the sequence SEQ ID No. 2 of the cDNA of the human SPG4 gene;
- the sequence SEQ ID No. 73, corresponding to a fragment of murine spastin encoded by the sequence SEQ ID No. 72 of the incomplete cDNA of the mouse Spg4 gene, the sequence SEQ ID No. 73 is represented in Figure 4A, "SPAST\_MOUSE" line;
- the sequences of polypeptides which are homologues and variants of the polypeptide of sequence SEQ ID No. 3 or SEQ ID No. 73; and
- the sequences of the fragments thereof of at least 8, 10, 15, 30 or 50 consecutive amino acids.

Also preferably, a subject

It should be understood that the invention does not relate to polypeptides in natural form, i.e. they are not taken in their environment. Specifically, the invention relates to the peptides which are obtained by purification from natural sources, or obtained by genetic recombination or by chemical synthesis, and which can therefore  
5 include unnatural amino acids. The production of a recombinant polypeptide, which can be carried out using one of the nucleotide sequences according to the invention, is particularly advantageous since it makes it possible to obtain an increased degree of purity of the desired polypeptide.

The term "homologous polypeptide" will be intended to refer to the polypeptides  
10 which have certain modifications with respect to the reference polypeptide, such as in particular one or more deletions or truncations, an extension, a chimeric fusion and/or one or more substitutions, and the amino acid sequence of which shows at least 80%, preferably 90% or 95%, identity after alignment, with the reference amino acid sequence.

15 The term "variant polypeptide" (or protein variant) will be intended to refer to the set of polypeptides encoded by the variant nucleic acid sequences as defined above.

In particular, the variant polypeptides will comprise any polypeptide which is encoded by the mutated genomic sequence of the SPG4 gene of sequence SEQ ID No. 1, and the amino acid sequence of which has at least one mutation corresponding  
20 in particular to a truncation, deletion, substitution and/or addition of amino acid residues with respect to the sequence SEQ ID No. 3. In the present case, the variant polypeptides having at least one mutation will be linked to the pathologies of AD-HSP type.

The term "variant polypeptide" will also be intended to refer to any polypeptide  
25 resulting from mutation of a splice site in the genomic nucleic acid sequence SEQ ID No. 1.

The invention also comprises the cloning and/or expression vectors containing a nucleic acid sequence according to the invention.

The vectors according to the invention, characterized in that they include the  
30 elements which allow the expression and/or the secretion of said sequences in a host cell, or a cellular addressing sequence, also form part of the invention.

The vectors characterized in that they include a promoter and/or regulator sequence according to the invention also

should be able to be maintained stably in the cell and can, optionally, have particular signals which specify secretion of the translated protein.

These various control signals are chosen as a function of the host cell used. To this effect, the nucleic acid sequences according to the invention can be inserted into  
5 vectors which replicate autonomously in the host chosen, or vectors which integrate in the host chosen.

Among the systems which replicate autonomously, use will preferably be made, as a function of the host cell, of the systems of plasmid or viral type, the viral vectors possibly in particular being adenoviruses (Perricaudet et al., 1992), retroviruses,  
10 lentiviruses, poxviruses or herpesviruses (Epstein et al., 1992). Those skilled in the art know the technology which can be used for each of these systems.

When integration of the sequence into the chromosomes of the host cell is desired, use may be made, for example, of the systems of plasmid or viral type; such viruses will, for example, be retroviruses (Temin, 1986), or AAVs (Carter, 1993).

15 Among the nonviral vectors, preference is given to naked polynucleotides such as naked DNA or naked RNA according to the technique developed by the company VICAL, yeast artificial chromosomes (YAC) for expression in yeast, mouse artificial chromosomes (MAC) for expression in murine cells and, preferably, human artificial chromosomes (HAC) for expression in human cells.

20 Such vectors will be prepared according to the methods commonly used by those skilled in the art, and the clones resulting therefrom can be introduced into a suitable host using standard methods, such as for example lipofection, electroporation or heat shock.

The invention also comprises the host cells, in particular the eukaryotic and  
25 prokaryotic cells, transformed with the vectors according to the invention, and also the transgenic animals, except humans, comprising one of said transformed cells according to the invention.

Among the cells which can be used for these purposes, mention may of course be made of bacterial cells (Olins and Lee, 1993), but also yeast cells (Buckholz, 1993),  
30 as well as animal cells, in particular



















The results of all these sequence analyses were visualized using the Genotator sequence annotation program.

2) cDNA cloning

5 The cDNA of the SPG4 gene was isolated through 5' and 3' RACE-PCR experiments on polyA<sup>+</sup> RNAs of foetal brain, adult brain and adult liver, using the Marathon cDNA amplification kit (Clontech) according to the supplier's instructions. A first PCR followed by an internal PCR were carried out with various pairs of primers, the sequences of which are indicated in Table 1 hereinafter:

**Table 1**  
**Primers used for the RACE-PCRs and the cDNA amplifications**

Primer	Sequence (5'-3')	5' position pair/PCR product size		
SPA_5RACE5	CGGAGCTCCTCTTGGCTGCCATG	nt 405		
SPA_5RACE6	AGAAGCGCTGGCAGAGCCACACGAAG	nt 372		
SPA_5RACE7	AAGGCGACCAAACGCAGCAGCGCGAAG	nt 331		
SPA_3RACE1	AGGAGCAAGCTGTGGAATGGTATAAG	nt 550		
SPA_3RACE2	TGGTTATGGCCAAGGACCGCTTACAAC	nt 689		
SPA_3RACE3	CAAACGGACGTCTATAATGACAGTAC	nt 747		
SPA_3RACE4	TTAGGAATGTGGACAGCAACCTTGC	nt 1075		
SPA_3RACE5	CTTCTCTGAGGCCTGAGTTGTTCAC	nt 1207		
SPA_3RACE6	TGCTAGAATGACTGATGG			



**Table 2**  
**PCR primers for amplifying and sequencing the exons**

Exon	Product size	PCR program	Primer	Sequence (5'-3')
1	1048 bp	0	gSPAex1c	GTGAGCCGAACTGCACATTG
			gSPAex1m	CAAAGTCGACAGCTACAGTGC
			gSPAex1d	GGAAGTGTAGTTGAGTGGGA
			gSPAex1n	AGATGAGGCTCCGACCTAC
2	624 bp	3	gSPAex2a	AATGCCACACTTGTAATCTC
			gSPAex2m	TGTGAATATATCATAATTTGGG
			gSPAex2b	TACAGCAGTTCTCATGATG
3	812 bp	1	gSPAex3a	GACCAAATTGGTGCATGCATG
			gSPAex3m	ACATTTCCAATACATCCCAC
4	379 bp	3	gSPAex4a	ATT

13	1361 bp	4	gSPAex13a	CAGATTCAAGAAGACAGATC
			gSPAex13m	GCAATAATTCACCACACTTG
			gSPAex13n	GGTAGTTCTTGTTTCTGCTC
14	985 bp	4	gSPAex14a	CAAGTGTGGTGAATTATTGC
			gSPAex14m	GAGCTGAAAAGTATTCAGC
			gSPAex14n	TGCAAAGGACATAGCCAGTG
15	1076 bp	1	gSPAex15a	AGCCTCTGGAGATAGTATGC
			gSPAex15m	CTAGAACAGGGGTCACAGTC
			gSPAex15n	TTGGACTTCTTAACTTC
16	1404 bp	4	gSPAex16a	GCAGTATGCAAGAAATTGAAC
			gSPAex16m	GGCCTGTAATTTTCTTCTG
			gSPAex16b	GTAAGTGAATAGATACATGTAG
17	445 bp	3	gSPAex17a	GTGTAGCAGATCAACATAG
			gSPAex17m	CATCTTCAAGTTTGGTGCAC



Genomics); the 5 remaining genes could only be identified by sequencing the candidate region. One of these 5 novel genes showed homology in 3' of its coding region, with the genes encoding the AAA protein family (16). More thorough sequence analyses showed that this gene, named SPG4 (or SPAST), was composed of 17 exons and extended over a region of approximately 90 kb, covered by two adjacent BAC clones, D and G (cf. fig. 1B). The first three predicted exons of this gene were identified in BAC D, by two of the four exon prediction programs used, GRAIL II and GENSCAN; they show strong homology with a mouse blastocyst EST, AA560327. The last 14 exons are found in BAC G. The protein sequence deduced from exons 7 to 17 is significantly homologous to a subclass of the AAA family, which includes the Yta6p (21), TBP6 (21) and End 13 yeast proteins, and also the SKD1 mouse protein (22).

Of the four exon prediction programs FGENEH appears to be the most reliable and the most powerful, enabling detection of most of the genes of this chromosomal region at 2p21-p22. This observation also applies to the SPG4 gene, for which 15 exons could be demonstrated using this program, while only 4, 9 or 11 exons could be located using the Genie, GRAIL II and GENSCAN programs, respectively. The genomic organization of this gene (fig. 1B) could subsequently be confirmed by determining the sequence of the SPG4 cDNA. The intron/exon junctions are represented on table 3 hereinafter: the exon size ranges from 41 bp (exon 16) to 1.410 kb (exon 17), that of the introns ranging from 140 bp (intron 11) to 23.247 kb (intron 1).

Table 3

## Intron/exon organization of the SPG4 gene

Exon/ intron	Exon size (bp)	Position on the cDNA	Splice acceptor site	Splice donor site	Intron size (bp)
1	540	1		TGAGAAAG/gtaactaggggctgg	23 247
2	87	541	atthttatthtaag/CAGGACAG	AGGACAAAG/gtaagattgtattgt	1 943
3	84	628	aatttttctttcag/GTGAACAG	ACTTCTAG/gtatcaattaatgtat	9 190
4	96	712	cttctcgttgcatag/AGAAGATG	CCAGTCAG/gtgggttaggtaaac	15 745
5	188	808	acttttccctgtcag/AAAGTGGA	CTCATAAAG/gtattctgggacagta	876
6	134	996	ttttgatacctthtaag/GGTACTCC	GTGGACAA/gtaagttttgccatct	283
7	94	1 130	aggctctgtttcttag/TGGAACAG	GGCCTGAG/gtaagaactttatatt	10 735
8	75				



individuals of these three families (fig. 3B); this mutation engenders the loss of exon 16, followed by a reading frame shift in the abnormal transcript. None of the healthy members, including husbands and wives, carry this mutation of the splice site. The identification of the same mutation in all the affected members of these three Swiss families demonstrates the existence of a common ancestor, which had probably been suggested by the study of the haplotypes.

Three point mutations, 1210C->G, 1468G->A and 1620C->T, which introduced amino acid substitutions into the protein sequence (S362C, C448Y and R499C), were respectively revealed by sequencing PCR fragments III and IV in the affected individuals of families 624, 4014 and 618. These three substitutions all involve a cysteine residue, inducing the loss or insertion of a cysteine in the protein sequence. A 1 bp deletion, 1520delT, which creates the appearance of a STOP codon inducing a truncated protein composed of 465 amino acids (aa), was detected in the affected individuals of family A. None of the five mutations summarized in table 4 hereinafter was found in the control individuals tested, whether they belong to the healthy siblings or to the spouses of the seven families analyzed herein. These five mutations significantly affect the protein sequence in a very conserved domain, or AAA cassette (23), which is composed of several protein motifs presumed to be responsible for the ATPase activity in all the members of the AAA family.

**Table 4**  
**Mutations in SPG4 in the patients suffering from AD-HSP**

Family	Location	Mutation <sup>a</sup>	Amino acid change <sup>b</sup>	Consequence
624	exon 7	1 210 C → G	S362C	missense
4 014	exon 11	1 468 G → A	C448Y	missense
A	exon 11	1 520 delT	466STOPcodon	nonsense
618	exon 13	1 620 C → T	R499C	missense
2 992	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift
5 226	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift
5 330	intron 15	1 813-2a → g	Δ aa564 → aa576 (PTC+7 aa)	loss of exon 16 + shift

32

<sup>a</sup> The nt positions refer to the sequence of the SPG4 cDNA.

<sup>b</sup> The aa positions refer to the spastin sequence.

The bases of the exons are indicated in upper case, those of the introns in lower case.

PTC+7 aa = "premature termination codon"

In addition to these five mutations described above, searches for heterozygous mutations, carried out on patients suffering from AD-HSP derived from 36 other families, made it possible to reveal 34 other mutations which modified or were likely to modify the product of expression of the SPG4 gene.

- 5        The characteristics of these 34 other mutations are summarized in table 5 hereinafter, into which the first five mutations mentioned above have also been inserted.









regulation, organelle biogenesis, i.e. control of transcription, etc. However, all these cellular mechanisms involve the assembly, the functioning or the degradation of protein complexes, which suggest that the members of the AAA family are so-called "chaperon" proteins.





- Köhler et Milstein. *Nature* 256, 495-497, 1975.
- Kwoh, D.Y. et al. (1989), *Proc. Natl. Acad. Sci. USA*, 86, 1173-1177.
- Landegren U., Kaiser R., Sanders J. & Hood L. *Science* 241: 1077-1080, 1988.
- Lizardi, P.M. et al. (1988), *Bio/technology*, 6, 1197-1202.
- 5 Luckow, V.A. (1993), *Curr. Op. Biotechnology* 4, 564-572.
- Matthews, J.A. et al. (1988), *Anal. Biochem*, 169: 1-25.
- Miele, E.A. et al. (1983), *J. Mol. Biol.*, 171: 281-295.
- Olins, P.O., and Lee, S.C. *Curr. Op. Biotechnology* 4: 520-525, 1993.
- Perricaudet, M., Stratford-Perricaudet, L. and Briand, P. *La Recherche* 23: 471-473, 1992.
- 10 Rohlmann, A., Gotthardt, M., Willnow, T.E., Hammer, R.E., and Herz, J. *Nature Biotech.* 14: 1562-1565, 1996.
- Rolfs, A. et al. (1991), Berlin: Springer-Verlag.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. *Molecular cloning: a laboratory manual*. Sec. Ed. Cold Spring Harbor Lab., Cold Spring Harbor, New York.
- 15 Segev, D., (1992), Kessler C. Springer Verlag, Berlin, New York, 197-205.
- Stone, B.B. et al. (1996). *Mol. and Cell. Probes*, 10: 359-370.
- Stewart J.M. et Yound J.D., *solid phase peptides synthesis*, Pierce Chem. Company, Rockford, 111, 2ème éd., (1984).
- Suggs S.V., Wallace R.B., Hirose T., Kawashima E.H. and Itakura K. *PNAS* 78: 6613-6617, 1981.
- 20 Temin, H.M. Retrovirus vectors for







32. Use of a nucleic acid sequence according to one of Claims 1 to 6, and 12, of a polypeptide according to one of Claims 13 to 15, and 25, of a vector according to either of Claims 16 and 17, of a cell according to Claim 18, of a mammal according to either of Claims 19 and 20 or of an antibody according to Claim 26, for studying the expression or the activity of the SPG4 gene.
- 5

**ORIGINAL**

CABINET REGIMBEAU  
INDUSTRIAL PROPERTY CONSULTANT

26, Avenue Kléber  
75116 PARIS  
[Illegible signature]







32. Use of a nucleic acid sequence according to one of Claims 1 to 6, and 12, of a polypeptide according to one of Claims 13 to 15, and 25, of a vector according to either of Claims 16 and 17, of a cell according to Claim 18, of a mammal according to either of Claims 19 and 20 or of an antibody according to Claim 26, for studying the  
5 expression or the activity of the SPG4 gene.



FIGURE 1A

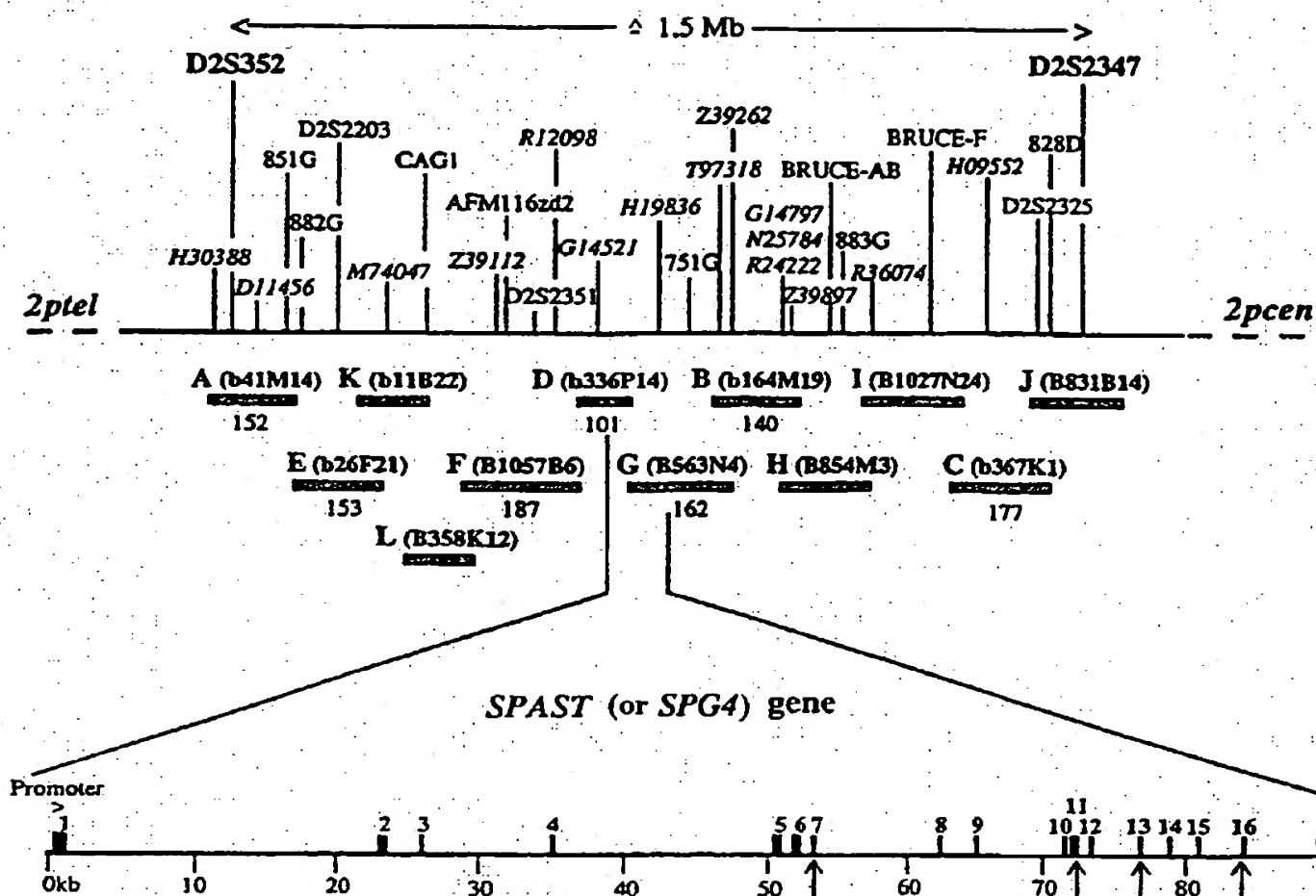
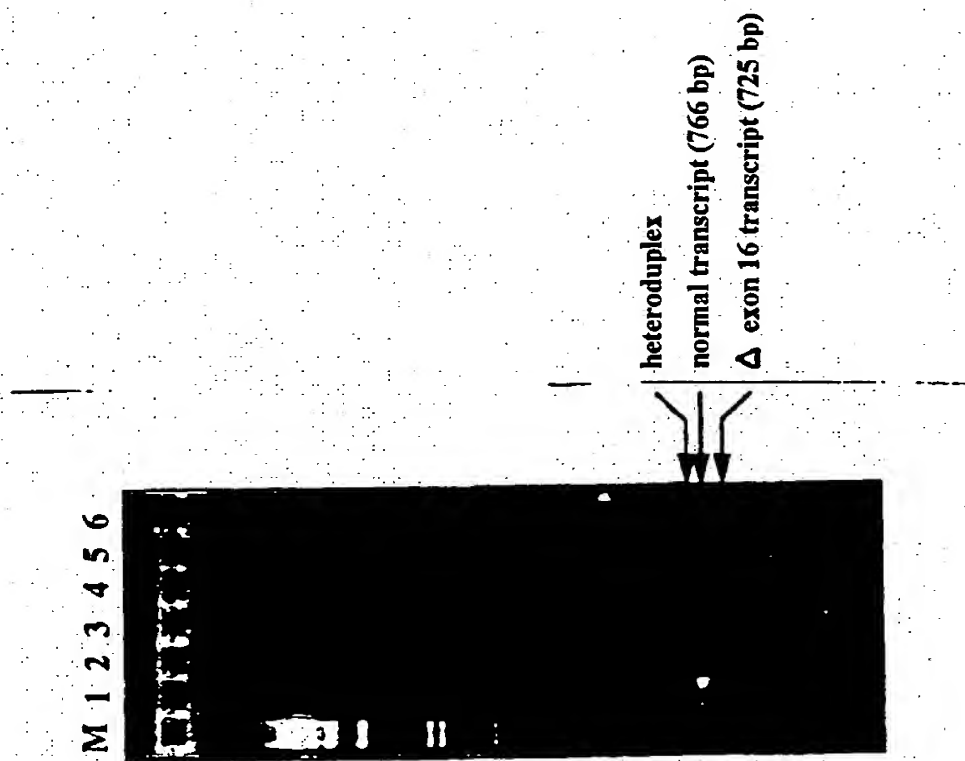


FIGURE 1B



A



B

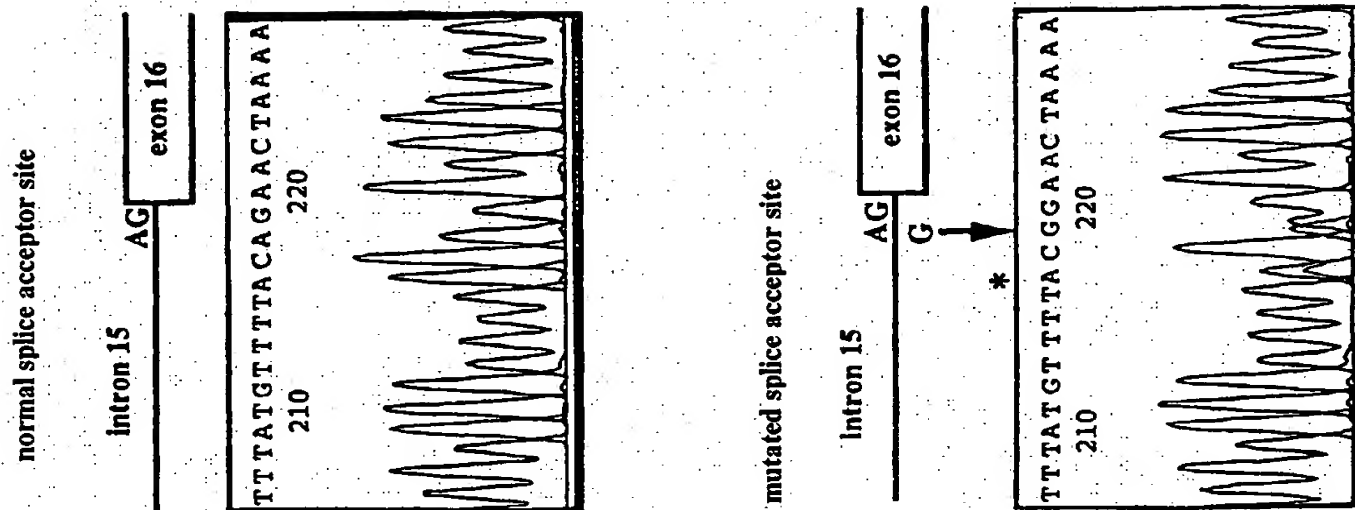


FIGURE 3







